A Cooperation Scheme for User Fairness and Performance Enhancement in NOMA-HCNm

A Cooperation Scheme for User Fairness and Performance Enhancement in NOMA-HCN

Pragya Swami, Vimal Bhatia, *Senior Member, IEEE*, Satyanarayana Vuppala, *Member, IEEE*, and Tharmalingam Ratnarajah, *Senior Member, IEEE*

***Abstract*—Rapid increase in number of cellular users and high demand for data has lead to the formation of multi-tier networks. Non-orthogonal multiple access (NOMA) has proved to be an efficient method to cater to the paradigm shift from 4G to 5G. This paper employs NOMA in a heterogeneous cellular network (HCN) consisting of a macro base station (MBS) tier underlaid with femto base station (FBS) tier and device to device (D2D) tier, where NOMA is employed in FBS and D2D tier only. The congestion at the MBS tier is relieved by offloading macro users (MU) to the FBS tier. The offloaded MU are further supported by the D2D tier when the FBS tier fails to find a corresponding pairing user for the incoming offloaded MU. Since, absence of pairing user means outage for offloaded MU, D2D cooperation is employed which decreases the rate outage probability by 86.87% for the MU offloaded as cell edge user (CEU) in comparison to no cooperation. Also, a 3 times increase in ergodic rate and 4 times increase in sum ergodic rate for MU offloaded as CEU is achieved using cooperation from D2D tier. Verification of the results is done using Monte Carlo simulations.**

***Index Terms*—Non-orthogonal multiple access, stochastic geometry, offloading, heterogeneous cellular network, D2D groups, cooperation**

I. INTRODUCTION

The present cellular network is unable to withstand the surge in both data demand and number of users. This challenge is met by utilizing low powered femto base stations (FBS) in the network thereby making it a heterogeneous cellular network (HCN) [1], [2], where FBS tiers employ orthogonal multiple access (OMA) technique. The FBS tier work in open, closed, or hybrid access mode [3], [4]. The research has shifted to 5G technologies and beyond owing to the challenges that could not be accommodated by the current 4G technologies. Non orthogonal multiple access (NOMA), an enabling technique for 5G, has proved to achieve higher spectral gains as compared to OMA [5] and is a viable solution for future dense networks and Internet of Things (IoT) devices. OMA is a well known multiple access technique adopted by the 4G communication systems [6], [7]. Performance of NOMA in downlink with randomly deployed users is studied in [8] which proves the superior ergodic rates achieved by NOMA over OMA. HCN and NOMA combined together are studied in [9]–[11]. Since NOMA supports multiple users at the same time and in the same frequency slots, hence, strong co-channel interference (CCI) affects all users served in that particular time which increases with increase in the number of users. Due to high CCI, large number of users cannot perform NOMA jointly. Hence, a hybrid multiple access system is formed where users are combined into groups (or pairs) that perform NOMA orthogonally with other groups (or pairs). In [12] the impact of user pairing in NOMA is studied and [13] studies a low complexity user pairing algorithm.

In [14], cooperation scheme for FBS is studied where FBS tier lowers its power to safeguard the other tier from cross-tier interference and in [15], cooperation techniques are studied from NOMA perspective. Authors in [16] analyses NOMA in a cognitive underlay network wherein the interference from primary network to secondary users is considered. NOMA involves splitting of power between users with different channel conditions. Thus, user fairness becomes an important issue in NOMA which is dealt in [17], where a dynamic user clustering problem is formulated from a fairness perspective and power allocation coefficients for the users in each cluster is optimized. In this direction, cell boundary users performance is also analyzed in [18]. Offloading in multi-tier environment, which involves handing some users to the less congested tier, is studied in [19]–[21]. Bypassing the base station (BS) is also considered as an option to reduce load on the MBS tier. Such a communication is termed as device to device (D2D) communication and is studied in detail in [22]–[24]. In [22] and [24], mode selection in underlay D2D network is studied, while [23] investigate an efficient way of reusing the downlink resources for cellular and D2D mode communication. A step further from D2D pairs, [25] studies D2D groups that use NOMA as their transmission technique to serve multiple D2D receivers. Unlike MBS, FBS is a device like wireless local area network router where it supports dense users with far varying needs like IoT devices and high definition video transmissions at the same time [26]. NOMA enables power splitting for multiple users with different needs. Hence, NOMA at FBS tier is more likely than at the MBS tier for dense networks and is studied in this work

A. Motivation and Contribution

Motivated by the advantage of HCN in meeting the explosive data demands, offloading for load balancing in multi-tier environment and NOMA in meeting 5G requirements we

This work is supported by the MeitY, DST UKIERI (DST/INT/UK/P-129/2016 and DST UKIERI-2016-17-0060), Indian Institute of Technology Indore, Indore, India, and University of Edinburgh, Edinburgh, U.K. This work of T. Ratnarajah is also supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P009549/1.

P. Swami and V. Bhatia are with Indian Institute of Technology Indore, Indore, India; S. Vuppala is with United Technologies Research Center, Cork, Ireland, and T. Ratnarajah is with the Institute for Digital Communications, University of Edinburgh, Edinburgh, U.K.

propose a network model combining NOMA and offloading in HCN. Two types of users are considered in this work namely; cell center users (CCU), which have good channel condition, and the second type of users are cell edge users (CEU), that suffers from poorer channel condition. The FBS tier employs NOMA and the congestion at MBS tier is managed by offloading the users to FBS tier. FBS tier pairs the incoming offloaded MU (OMU) with an appropriate/corresponding user (i.e., a CEU is paired with a CCU and vice versa), called as pairing user (PU) and performs NOMA on the pair. However, it is not always possible for an FBS to find a corresponding pair for the OMU. In such a case, D2D tier cooperates to serve the OMU using NOMA, which is explained in detail in Section II-E. Useful observations are made from offloading users to FBS tier with NOMA and cooperation from D2D tier. The metric used for the analysis are namely, rate outage probability [21] and ergodic rate.

Primary contributions of this work are listed below:

- An analytical framework is designed for an HCN with NOMA (HCN-NOMA) network. The congestion at MBS tier is handled by offloading MUs to FBS tier. Since FBS uses NOMA, a check is performed on the OMU to ascertain whether the user is accommodated as a CEU or CCU with respect to the available PU at the FBS.
- Two different cases of offloading are considered. Case I assumes presence of a PU with FBS and the incoming OMU's channel condition is compared with the available PU to find whether it will be served as a CCU or CEU with respect to the available PU. In Case II, we search for a corresponding PU depending on whether the incoming OMU is a CCU or CEU.
- Performance analysis of the OMU is done in terms of rate outage probability. For Case I, the performance depends on whether the OMU is a CCU or CEU with respect to the available PU, and also on the difference in channel gain between the OMU and the PU. For Case II, the performance depends on the availability of a corresponding PU. Some useful observations on performance enhancement and user fairness are drawn.
- For Case II, since, it may not always be possible for FBS to find an appropriate corresponding PU for the OMU, we discuss the case when a corresponding PU is unavailable for the OMU (for instance, when both the OMU and available PU are CEU). Such a scenario is dealt by using cooperation from the D2D tier.
- Ergodic rate for the offloaded CCU and CEU is also calculated. Ergodic rate at the OMU and sum ergodic rate at FBS (with NOMA) is compared for before and after cooperation.

Rest of the paper is organized as follows. System model is given in Section II. Section III derives some useful expressions for rate outage probability and ergodic rate for the three considered tiers and the OMU. Numerical results are discussed in Section IV. Finally, the work is concluded in Section V.

II. System Model

A three-tier network comprising of MBS, FBS, and D2D transmitters (DT) is considered for the analysis. The FBS and D2D tier are assumed to be underlaid with the MBS tier and work under open access mode. NOMA can be employed in FBS tier and D2D tier. When NOMA is used, the power coefficients are denoted as a_n for the n^{th} strongest user in NOMA. The spatial distribution of nodes follow independent Poisson point process (PPP). For the t^{th} tier this distribution is denoted by Ω_t with density λ_t where $t \in \{m, f, d\}$ for MBS tier, FBS tier, and D2D tier, respectively. Users are distributed according to an independent PPP Ω_u with density λ_u. The transmit power of tier t is denoted by P_t and $\mathcal{Y}_t$ denotes the coverage range of t^{th} tier. Each DT has its intended receiver uniformly distributed in the proximity range (PR) of DT which is defined as the maximum distance/radius upto which a DT is capable of D2D communication. A DT may or may not use NOMA depending upon the number of D2D receivers (DRs) it needs to serve. When a DT has a single DR, it is called as a D2D pair and the DT does not use NOMA. However, when the number of DRs are greater than one, the DT along with all its DRs is referred to as a D2D group [25], and the DT uses NOMA to serve its DRs. Since, we use PPP model to distribute the DTs, we assume sufficiently high λ_d, which implies a large mean ($\lambda_d \times |A|$) for Poisson random variable, where $|A|$ denotes the area within which the DTs are distributed, resulting in large number of DTs [21], [27]. Hence, we assume that a DT will always be available for the OMU to perform D2D cooperation, whenever required (as explained in Section II-E).

An MU connects to the MBS according to the nearest neighbor (NN) connection policy. Bounded path loss model is considered as $P(r) = \frac{1}{1+r_t^{\nu_t}}$, to ensure that the path loss is always smaller than one even for small distances [28], where ν_t is the path loss exponent for t^{th} tier and r_t represents the distance between typical user and tagged BS (BS that serves the typical user) of the t^{th} tier, respectively. Hence, the total channel gain for the typical user of t^{th} tier is given by $|h_t|^2 = |\hat{h}_t|^2 P(r)$, where $\hat{h}_t$ is assumed to follow Rayleigh distribution. The overall system transmission bandwidth is assumed to be 1 Hz. R is the target data rate of a typical user, assumed to be same for all the tiers. Congestion at the MBS tier is relieved by offloading users to FBS tier. Furthermore, cooperation from D2D tier is proposed when the FBS tier fails to serve the OMU, as explained later in this paper. The communication regarding D2D cooperation is managed between the FBS and DTs by feedbacks using localization techniques [29], [30] to share the necessary information. For a successful communication, i.e., for a non-outage condition, the instantaneous rate at the user should be greater than a target data rate of the user.

For the analysis we consider two distinct cases, namely

- Case I: We assume that a PU is always available at the FBS and the incoming OMU is paired with this PU either as CCU (Fig. 1, Case I (a)) or CEU (Fig. 1, Case I (b)), depending on its channel condition with respect to the PU. This pair is then served using NOMA.
- Case II: Unlike Case I, we assume the offloaded MU to be either a CEU or CCU and search for a corresponding PU accordingly. When FBS fails to find a corresponding

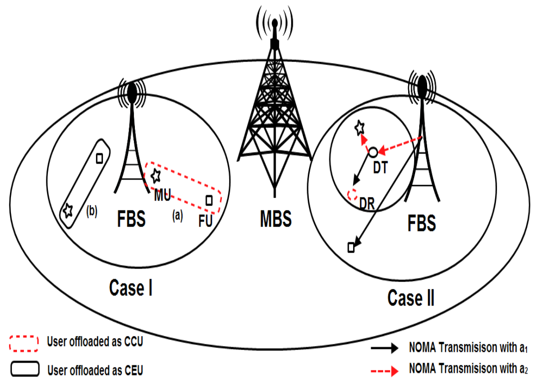


Fig. 1: System Model

PU for the OMU, D2D tier cooperates to serve the OMU as shown in Fig. 1, Case II. Cooperation from D2D tier is discussed in detail in Section II-E, when both the OMU and available PU are CEU.

***Note*:** Throughout the paper, m will represent parameters for MBS tier, f for FBS tier, and d for D2D tier.

A. Signal to Interference and Noise Ratio at Typical MU

The signal intended for a typical MU is given by x_m. The signal transmitted by the MBS can be written as $X_{m,tx} = \sqrt{P_m} x_m$ and the received signal[1] can be written as $X_{m,rx} = \sqrt{P_m} x_m \tilde{h}_m + n_m$, where n_j denotes channel noise at user of j^{th} tier such that $j \in \{m, f, d\}$.

Given a signal X, the useful signal power, noise and/or interference power can be easily calculated using $P = \mathbb{E}[XX^*]$, where $\mathbb{E}[.]$ denotes the statistical expectation. Hence, the signal to interference and noise ratio (SINR) at the typical MU can be written as

$$\text{SINR}_m = \frac{P_m \rho_m |\tilde{h}_m|^2}{\sum_t \rho_t^I \mathcal{I}_t + 1}, \tag{1}$$

where $\rho_m = \mathbb{E}\left[x_m^2\right]/\sigma_m^2$ and σ_j^2 denotes noise variance such that $j \in \{m, f, d\}$. ρ_t^I denotes the transmit SNR from BS of t^{th} tier responsible for interference and is given by $\rho_t^I = P_t/\sigma_m^2$. $\rho_t^I \mathcal{I}_t$ denotes the interference from t^{th} tier. Without loss of generality, assuming the typical MU to be located at the origin according to the Slivnyak's theorem [31] and the tagged MBS at m_0, the co-tier interference from MBS tier to the typical MU is written as $\rho_m^I \mathcal{I}_m$ such that $\mathcal{I}_m = \sum_{i \in \Omega_m/\{m_0\}} |\tilde{h}_i|^2$, where $|\tilde{h}_i|^2$ denotes the total channel gain from i^{th} MBS to the typical MU, and $\mathcal{I}_{t_m} = \sum_{i \in \Omega_{t_m}} |\tilde{h}_j|^2$, where $t_m \in t/\{m\}$ are the tiers contributing to cross-tier interference at the typical MU, $|\tilde{h}_j|^2$ denotes the total channel gain from j^{th} transmitter of the t_m^{th} tier to the typical MU.

[1]Throughout the paper, the use of hat, $\hat{h}$, denotes Rayleigh distribution, use of tilde, $\tilde{h}^2$, denotes that the channel gain are unordered, and h^2 denotes ordered channel gain.

B. SINR at Typical Femto User (without NOMA)

Similar to Section II-A, we may directly write the SINR at a typical femto user (FU) as

$$\text{SINR}_f = \frac{\rho_f P_f |\tilde{h}_f|^2}{\sum_t \rho_t^I \mathcal{I}_t + 1}, \tag{2}$$

where ρ_f denotes the transmit SNR at FBS and is given by $\rho_f = \mathbb{E}\left[x_f^2\right]/\sigma_f^2$, where x_f denotes the intended message signal for the typical FU. $\rho_t^I = P_t/\sigma_f^2$ denotes the transmit SNR from the transmitter of t^{th} tier responsible for interference at typical FU. Without loss of generality, assuming the typical FU to be located at the origin according to the Slivnyak's theorem [31] and the tagged FBS at f_0, the co-tier interference from FBS tier to typical FU is written as $\rho_f^I \mathcal{I}_f$ such that $\mathcal{I}_f = \sum_{i \in \Omega_f/\{f_0\}} |\tilde{h}_i|^2$, where $|\tilde{h}_i|^2$ denotes the total channel gain from i^{th} FBS to the typical FU, and $\mathcal{I}_{t_f} = \sum_{i \in \Omega_{t_f}} |\tilde{h}_j|^2$, where $t_f \in t/\{f\}$ are the tiers contributing to cross-tier interference at the typical FU, $|\tilde{h}_j|^2$ denotes the total channel gain from j^{th} transmitter of the t_f^{th} tier to the typical FU.

C. SINR at Typical FU (with NOMA)

Let us assume that M_f users are being served by an FBS using NOMA, and the channel gains of the M_f users are ordered as $|h_1^f|^2 \leq \ldots \leq |h_{M_f}^f|^2$. Based on this ordering of the users' channel gain, NOMA orders the respective power allocation factors as $a_1 \geq \ldots \geq a_{M_f}$. Given x_i as the intended signal for i^{th} user, $\mathbb{E}[x_i^2]$ is assumed to be equal $\forall i \in (1, 2, \ldots, M_f)$. The signal transmitted by the FBS is given by $X_{f,tx} = \sum_{i=1}^{M_f} x_i \sqrt{a_i P_f}$. Hence, the signal received by user k is given by $X_{f,rx} = h_k^f (\sum_{i=1}^{M_f} x_i \sqrt{a_i P_f}) + n_f$.

SINR at user k to decode message of user j ($j < k$) is given as [16]

$$\text{SINR}_{k \to j}^f = \frac{\rho_f P_f a_j |h_k^f|^2}{\rho_f P_f |h_k^f|^2 \sum_{l=j+1}^{M_f} a_l + \sum_t \rho_t^I \mathcal{I}_t + 1}, \tag{3}$$

where $\rho_f = \mathbb{E}[x_i^2]/\sigma_f^2$ denotes the transmit SNR at FB, a_n denotes power allocation factor for user with index $n = \{k, j, l\}$. $\rho_t^I = P_t/\sigma_f^2$ denotes the transmit SNR from the transmitter of t^{th} tier responsible for interference. Without loss of generality, assuming the k^{th} typical FU to be located at the origin according to the Slivnyak's theorem [31] and the tagged FBS at f_0, the co-tier interference from FBS tier to k^{th} typical FU is written as $\rho_f^I \mathcal{I}_f$ such that $\mathcal{I}_f = \sum_{i \in \Omega_f/\{f_0\}} |\tilde{h}_i|^2$, where $|\tilde{h}_i|^2$ denotes the total channel gain from i^{th} FBS to the typical FU, and $\mathcal{I}_{t_f} = \sum_{i \in \Omega_{t_f}} |\tilde{h}_j|^2$, where $t_f \in t/\{f\}$ are the tiers contributing to cross-tier interference at the k^{th} typical FU, $|\tilde{h}_i|^2$ denotes the total channel gain from i^{th} transmitter of the t_f^{th} tier to the k^{th} typical FU. SINR at user k to decode its own message is given by

$$\text{SINR}_k^f = \frac{\rho_f P_f a_k |h_k^f|^2}{\rho_f P_f |h_k^f|^2 \sum_{l=k+1}^{M_f} a_l + \sum_t \rho_t^I \mathcal{I}_t + 1}. \tag{4}$$

D. SINR at typical DR of D2D group (with NOMA)

Let us assume that there are M_d number of DRs in a D2D group. The channel gains and power allocation coefficients of the M_d DRs are assumed to follow similar order as for the FBS (with NOMA) tier and are given as $|h_1^d|^2 \leq \ldots \leq |h_{M_d}^d|^2$, and $a_1 \geq \ldots \geq a_{M_d}$, respectively. Similar to Section II-C, SINR at k^{th} typical DR to decode message of j^{th} DR ($j < k$) is given by

$$\text{SINR}_{k\to j}^d = \frac{\rho_d P_d a_j |h_k^d|^2}{\rho_d P_d |h_k^d|^2 \sum_{l=j+1}^{M_d} a_l + \sum_t \rho_t^I \mathcal{I}_t + 1}, \tag{5}$$

where $\rho_d = \mathbb{E}[x_i^2]/\sigma_d^2$ denotes the transmit SNR at DT, and a_n denotes power allocation factor for n^{th} DR of D2D group. Without loss of generality, assuming the k^{th} typical DR to be located at the origin according to the Slivnyak's theorem [31] and the tagged DT at d_0, the co-tier interference from D2D tier to the k^{th} typical DR is written as 4 $\rho_d^I \mathcal{I}_d$ such that $\mathcal{I}_d = \sum_{i\in\Omega_d/\{d_0\}} |\tilde{h}_i|^2$, where $|\tilde{h}_i|^2$ denotes the total channel gain from i^{th} DT to the typical DR, and $\mathcal{I}_{t_d} = \sum_{i\in\Omega_{t_d}} |\tilde{h}_j|^2$, where $t_d \in t/\{d\}$ are the tiers contributing to cross-tier interference at the k^{th} typical DR, $|\tilde{h}_j|^2$ denotes the total channel gain from j^{th} transmitter of the t_d^{th} tier to the k^{th} typical DR. SINR at the k^{th} typical DR to decode its own message is given by

$$\text{SINR}_k^d = \frac{\rho_d P_d a_k |h_k^d|^2}{\rho_d P_d |h_k^d|^2 \sum_{l=k+1}^{M_d} a_l + \sum_t \rho_t^I \mathcal{I}_t + 1}. \tag{6}$$

E. Cooperation from D2D Tier

In this section, we consider the case when the OMU is a CEU and FBS cannot find a corresponding CCU for pairing. For instance, instead of a CCU, the available PU at FBS is also a CEU. This means that both the OMU and the PU have nearly similar channel conditions and hence, FBS cannot perform NOMA on the OMU. NOMA in D2D tier (more specifically in D2D groups) enables us to deal with such a situation. We have designed an analytical framework with D2D pairs, which transforms into D2D groups to cooperate for serving the OMU and the PU with similar channel conditions using NOMA. When the MU offloaded as a CEU falls under outage due to unavailability of a CCU for pairing, FBS serves this OMU and the available PU, which have nearly similar channel conditions (both CEU in our case), with cooperation from D2D tier. For this, FBS first serves one CEU (say the CEU available other then the offloaded CEU) and a DT that can act as a CCU for FBS as shown in Fig. 1, Case II. DT then serves the offloaded CEU and its DR by forming a D2D group and using NOMA in the group, i.e., D2D pair is transformed to a D2D group. The DT performing D2D cooperation is termed as cooperating DT. Whether the offloaded user served using D2D cooperation is considered as a CCU or CEU by the cooperating DT is decided using NOMA compatibility probability as discussed in detail in Section III-B2.

Note: D2D cooperation takes place only when the MU is offloaded as a CEU and the corresponding CCU is unavailable for pairing. D2D cooperation is not performed when the offloaded MU is a CCU and the corresponding CEU is unavailable for pairing. This is because when the OMU is a CCU, and the available PU is also a CCU, according to the proposed D2D cooperation, as given in Section II-E, FBS would require a DT that can act as a CEU. A DT lying between the OMU and the FBS cannot act as a CEU in this case. Hence, the DT that can get paired with the available PU and perform D2D cooperation will be lying beyond the OMU. This would involve sending the desired signal from the FBS to a DT far from the offloaded CCU, which will then send the desired message back to the offloaded CCU. This would involve transmission of the desired signal to a longer distance as is required and hence would involve unnecessary power wastage.

III. Performance Analysis

In this section, rate outage probability (or outage probability) and ergodic rate expressions are derived for the MBS tier, FBS tier, and D2D tier using stochastic geometry. Furthermore, offloading probability, NOMA compatibility (NC) probability, corresponding PU probability, total outage probability, and total ergodic rate for the considered system are also calculated. A non-outage condition is satisfied when the instantaneous rate is higher than the target data rate of the user [21].

A. Rate Outage Analysis

1) Rate Outage Analysis for MBS Tier: The rate outage probability of a typical MU is given as follows.

Proposition 1: Conditioned on the fact that MU connects to nearest MBS and small scale fading is Rayleigh distributed, the outage probability of a typical MU is given as

$$\mathcal{P}_O^m = \pi\lambda_m \mathcal{Y}_m^2 \sum_{n=0}^{N} b_n^m e^{-c_n^m \frac{\phi}{\rho_m P_m}} \prod_t \mathcal{L}_{\mathcal{I}_t}(s_m \rho_t^I), \tag{7}$$

where N is a parameter to ensure a complexity-accuracy trade-off, $b_n^t = -w_N \sqrt{1-\theta_n^2}\left(\frac{1}{2}(\theta_n+1)\right) e^{-\pi\lambda_t\left(\frac{1}{2}(\theta_n+1)\mathcal{Y}_t\right)^2}$ such that $t \in \{m, f, d\}$, $b_0 = -\sum_{n=1}^{N} b_n^t$, $c_n^t = 1 + \left(\frac{\mathcal{Y}_t}{2}\theta_n + \frac{\mathcal{Y}_t}{2}\right)_t^{\nu}$, $c_0 = 0$, $w_N = \frac{\pi}{N}$, $\theta_n = \cos\left(\frac{2n-1}{2N}\pi\right)$ [8]. The $\phi = 2^{2R} - 1$ denotes the SINR threshold, and $s_t = \frac{c_n^t \phi}{\rho_t P_t}$. $\mathcal{L}_{\mathcal{I}_t}(s)$ is the Laplace transform (LT) of interference from t^{th} tier and is calculated [28] as

$$\mathcal{L}_{\mathcal{I}_t}(s) = e^{\pi\lambda_t\left(s^{\delta_t}\Gamma(1-\delta_t, s) - s^{\delta_t}\Gamma(1-\delta_t)\right)}, \tag{8}$$

where $\delta_t = 2/\nu_t$, $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}$ and $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}$.

Proof: Please see Appendix A.

2) Rate Outage Analysis for FBS Tier (without NOMA): Similar to Proposition 1, the rate outage probability of typical FU conditioned on the uniform distance from the FBS is given as follows.

Proposition 2: Conditioned on the uniform distance from FBS, the outage probability at typical FU is given as

$$\mathcal{P}_O^f = \frac{1}{\mathcal{Y}_f} \sum_{n=0}^{N} b_n^f e^{-c_n^f \frac{\phi}{\rho_f P_f}} \prod_t \mathcal{L}_{\mathcal{I}_t}(\rho_t^I s_f). \tag{9}$$

Proof: Please see Appendix B.

3) Rate Outage Analysis for FBS Tier (with NOMA): The rate outage probability at the k^{th} typical FU is expressed as follows.

Proposition 3: Conditioned on the uniform distance of a typical FU from FBS and ordered channel gain of the FUs, the rate outage probability at k^{th} typical FU is given as

$$\mathcal{P}_k^f = \psi_k^f \sum_{r=0}^{M_f-k} \binom{M_f-k}{r} \frac{(-1)^r}{k+r} \sum_{T_k^r} \binom{k+r}{q_0 \dots q_N} \left(\prod_{n=0}^{N} (b_n^f)^{q_n}\right) e^{-\sum_{n=0}^{N} q_n c_n^f \frac{\epsilon_{max}^f}{\rho_f P_f}} \prod_t \mathcal{L}_{\mathcal{I}_t}(s_f \rho_t^I), \quad (10)$$

where $\epsilon_{max}^t = \max(\epsilon_1^t, \epsilon_2^t, \dots, \epsilon_k^t)$ such that $t \in \{f, d\}$ and ϵ_j^t is calculated as $\epsilon_j^t = \phi_j/(a_j - \phi_j \sum_{i=j+1}^{M_t} a_i)$, where $\phi_j = 2^{R_j} - 1$ and R_j denotes the target data rate of j^{th} user. $R_j = R \ \forall j \in (1, 2, \dots, M_t)$, $\psi_k^t = \frac{M_t!}{(k-1)!(M_t-k)!}$, $s_f = \frac{\epsilon_{max}^f \sum_{n=0}^{N} q_n c_n^f}{\rho_f P_f}$, $T_k^r = \left(q_0, \dots, q_N \mid \sum_{i=0}^{N} q_i = k+r\right)$, $\binom{k+r}{q_0 \dots q_N} = \frac{M_f!}{q_0! \dots q_N!}$.

Remark 1: It can be noted from (10) that the user with $k = 1$ does not perform SIC, hence the term ϵ_{max}^f equals ϵ_1^f. Also the outage probability in (10) is dependent on the transmit SNR of MBS, FBS and D2D tier, and on the user's target rate. The dependence is directly proportional to the target rate and the transmit SNR of MBS and D2D tier, while it is inversely proportional to the transmit SNR of FBS tier as observed in Fig. 2.

Proof: Please see Appendix C.

4) Rate Outage Analysis for D2D tier (with NOMA): Similar to Section III-A3, Proposition 3, the expression for the outage probability at typical DR of D2D group may be written as

$$\mathcal{P}_k^d = \psi_k^d \sum_{r=0}^{M_d-k} \binom{M_d-k}{r} \frac{(-1)^r}{k+r} \sum_{T_k^r} \binom{k+r}{q_0 \dots q_N} \left(\prod_{n=0}^{N} {b_n^d}^{q_n}\right) e^{-\sum_{n=0}^{N} q_n c_n^d \frac{\epsilon_{max}^d}{\rho_d P_d}} \prod_t \mathcal{L}_{\mathcal{I}_t}(s_d \rho_t^I), \quad (11)$$

where $s_d = \frac{\epsilon_{max}^d \sum_{n=0}^{N} q_n c_n^d}{\rho_d P_d}$, $\binom{k+r}{q_0 \dots q_N} = \frac{M_d!}{q_0! \dots q_N!}$.

B. Offloading, NOMA Compatibility, and Corresponding PU Probability

This section discusses the offloading probability, NOMA compatibility (NC), and corresponding PU probability. The offloading probability of an MU being offloaded from MBS to FBS tier is conditioned on the long term averaged biased-received-power (BRP) received from the FBS and MBS. The NC probability describes whether the OMU is a CEU or CCU with respect to the available PU at the FBS. When the OMU is assumed to be offloaded as a CEU, a corresponding CCU is searched for pairing and vice versa. The corresponding PU probability is used to find whether corresponding PU for OMU is available or not. Case I uses NC probability, while corresponding PU probability is used in Case II.

1) Offloading Probability: Offloading probability from MBS tier to FBS tier can be calculated as follows.

Proposition 4: Offloading is based on maximum BRP [19], where a user is associated with the strongest BS in terms of long-term averaged BRP at the user. The closed form expression for the offloading probability for $\nu_m = 3$ and $\nu_f = 4$ is given as

$$\mathcal{P}^{m \to f} = -\frac{3}{8} \mathrm{E}\left(\frac{1}{4}, 2\pi\lambda_m \left(\frac{B_m P_m}{B_f P_f}\right)^{\frac{1}{2}} \mathcal{Y}_f^{8/3}\right) + \frac{3\Gamma\left(\frac{3}{4}\right)}{8(2\pi)^{3/4} \mathcal{Y}_f^2 \left(\lambda_m \left(\frac{B_m P_m}{B_f P_f}\right)^{\frac{1}{2}}\right)^{3/4}} - \frac{1}{2} e^{-2\pi\lambda_m \mathcal{Y}_m^2}, \quad (12)$$

where B_m and B_f are the bias factor for MBS and FBS tier respectively. $\mathrm{E}(n, x)$ evaluates the exponential integral as $\mathrm{E}(n, x) = \int_1^\infty e^{-xt}/t^n \, dt$ and $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$ is the complete gamma function.

Proof: Please see Appendix D.

2) NOMA Compatibility Probability: When an MU is offloaded to an FBS, it is necessary to find out whether or not the user can be accommodated within the NOMA norms of different channel conditions. The probability of whether FBS can apply NOMA to the OMU or not is decided on whether the OMU satisfies the sufficiently different channel condition criterion and whether it will be accommodated as a CCU or a CEU. This condition for the OMU is checked with respect to the available PU. Assuming that index k stands for the OMU and n for the available PU at the FBS, the probability of the OMU to be offloaded as a CCU with respect to the PU can be calculated as $\mathcal{P}_{NC} = \mathcal{P}\left(\frac{|h_n|^2}{|h_k|^2} < p\right)$, where p (satisfying $0 < p < 1$) represents the ratio of channel gain of the PU and the OMU. The probability density function (PDF) of ratio of two order statistics [32] is given as

$$f_{\frac{h_n^2}{h_k^2}}(z) = \frac{M_f!}{(n-1)!(-n+k-1)! \ (M_f-k)!} \sum_{j_1=0}^{(n-1)} \sum_{j_2=0}^{(-n+k-1)} \frac{(-1)^{j_1+j_2} \binom{n-1}{j_1} \binom{-n+k-1}{j_2}}{(z \ t_1 + t_2)^2}, \quad (13)$$

where $t_1 = j_1 - j_2 + k - n$, and $t_2 = M_f - k + 1 + j_2$. Hence, the probability can be calculated using $\mathcal{P}_{NC} = \int_0^p f_{(h_n^2/h_k^2)}(z) dz$. NC probability helps us differentiate whether the OMU is a CCU or CEU with respect to the available PU. The value of p signifies the amount of difference in the channel gains between the OMU and PU. Hence, results for different values of p are discussed in Section IV. A lower value of p signifies a large difference in the users' channel gain, while a large value of p signifies smaller difference in the users' channel gain.

Remark 2: A tractable analysis is done with $M_f = 2, k = 2$ (OMU), and $n = 1$ (PU). Hence, we get the NC probability as $\mathcal{P}_{NC} = 2p/(p+1)$. Also, the probability of OMU to be offloaded as a CEU with respect to the PU would require $p > 1$. Using simple mathematics, we can straightforwardly write the probability of OMU to be offloaded as a CEU with respect to the PU as $1 - \mathcal{P}_{NC}$.

3) Corresponding PU Probability: For Case I, we assume availability of PU with FBS and the incoming OMU is paired with the PU as CEU or CCU, depending on its channel condition with respect to that of the available PU. Unlike Case I, for Case II, when an OMU is received at FBS it searches for a corresponding PU depending on whether MU is offloaded as a CCU or CEU. This means that availability of a corresponding user is uncertain, and hence there might be a case that for an offloaded CEU we get another CEU for pairing due to the unavailability of CCU and vice versa. Hence, we derive the probability for presence of a corresponding CCU for pairing with the MU offloaded as a CEU and probability for presence of a corresponding CEU for pairing with the MU offloaded as a CCU.

We define the CCU region in this paper based on the equal strength boundary (ESB) concept [3], [33]. [3] defines a FBS coverage as the area inside which the average received power from the FBS is stronger than that from the MBS nearest to the FBS. The boundary of this region is termed as ESB, at which the average received power from the FBS equals the average received power from the MBS. The MBS considered in the calculation of ESB region is the one closest to the FBS around which ESB is to be calculated. The proof in [3] shows that ESB is a circular region where FBS is located at the center of ESB, and the ESB around the FBS increases linearly with the distance of the nearest MBS. The ESB is given as $\mathcal{Y}_{eq} = \left(\frac{P_f}{P_m}\right)^{\frac{1}{\nu}} r_c$, where r_c is the distance from FBS to its nearest MBS (assumed to be at origin). FBS has a higher signal strength, as compared to the signal strength of the nearest MBS, within the ESB while beyond this boundary MBS's signal strength dominates. For the sake of simplicity, we have not considered the effect from the transmissions of the underlaid DTs on the ESB. Hence, ESB is an outcome of the simultaneous transmission from MBS and FBS, only. To distinguish the CCU and CEU users, we use the concept of ESB. Since, the received power from FBS is stronger inside the ESB, hence, the users lying inside the ESB can be considered as users with good channel condition, therefore, are termed as CCU. Outside the ESB, the received signal strength from FBS is lower as compared to the nearest MBS, rendering poor channel condition for users beyond ESB, therefore, are termed as CEU. Hence, ESB is the CCU region and the CEU region is the annulus region between the FBS coverage and the ESB.

Proposition 5: Considering the ESB region to demarcate the CCU region, and the annulus region of the ESB and FBS coverage to mark the CEU region, the probability of finding a PU in CCU region ($\mathcal{P}_C^u$), and in CEU region ($\mathcal{P}_E^u$) can be expressed as

$$\mathcal{P}_C^u = 1 - \frac{\left(1 - e^{\pi(-C)\mathcal{Y}_m^2}\right)\lambda_m}{C}, \tag{14}$$

$$\mathcal{P}_E^u = 1 - \frac{\left(1 - e^{\pi(-C)\mathcal{Y}_m^2}\right)e^{-\lambda_u \pi \mathcal{Y}_f^2}\lambda_m}{C}, \tag{15}$$

where $C = \lambda_u \left(\frac{P_f}{P_m}\right)^{\frac{2}{\nu}} + \lambda_m$.

Remark 3: The corresponding PU probability is dependent on the density of both, the user as well as of the MBS tier. The dependence on user density is straightforward and we may directly say that as the user density increases the probability of finding a CCU or CEU for pairing will also increase. The dependence on MBS density is contributed to the formation of ESB around FBS. As the MBS density increases, the ESB decreases and hence, the probability of finding an CCU for pairing decreases while the probability of finding a CEU increases due to the wider annulus region with the decreased ESB. The decrease in ESB with the increase in MBS density is contributed to the fact that higher density of MBS tier would indicate that the nearest MBS to the FBS would lie closer as compared to the nearest MBS when density of MBS is low.

Proof: Please see Appendix E.

C. Total Outage Probability

In this section, we calculate the total outage probability when offloading and cooperation is performed.

1) Total Outage Probability after Offloading: Combining outage probability, offloading probability and NC probability, the total outage probability when a PU is assumed to be available with FBS for the OMU, i.e., for Case I, is given as

- MU is offloaded to FBS (without NOMA)

$$\mathcal{P}_T = (1 - \mathcal{P}^{m\to f})\mathcal{P}_O^m + \mathcal{P}^{m\to f}\mathcal{P}_O^f. \tag{16}$$

- MU is offloaded as a CCU at FBS (with NOMA)

$$\mathcal{P}_T^C = (1 - \mathcal{P}^{m\to f})\mathcal{P}_O^m + \mathcal{P}^{m\to f}\mathcal{P}_{NC}\mathcal{P}_k^f. \tag{17}$$

- MU is offloaded as a CEU at FBS (with NOMA)

$$\mathcal{P}_T^E = (1 - \mathcal{P}^{m\to f})\mathcal{P}_O^m + \mathcal{P}^{m\to f}(1 - \mathcal{P}_{NC})\mathcal{P}_k^f. \tag{18}$$

Remark 4: The above equations in (16), (17), and (18) describes the outage probability for Case I, and combine two situations, one where no offloading takes place (denoted by the first terms), and second when offloading occurs (denoted by the second terms). Equations (17) and (18) also includes the NC probability in their second terms as a check for whether the incoming OMU is a CEU or CCU with respect to the available PU. As can be observed from Fig. 3, the NC probability (or we may say that the value of p) has a huge impact on the outage probability of the OMU.

Similarly, we combine the outage probability, offloading probability and corresponding PU probability to calculate the total outage probability when a corresponding PU is searched by the FBS for the OMU based on whether OMU is a CEU or CCU, i.e., for Case II. Hence, the total probability can be written as

- The MU is offloaded as a CCU and a CEU is searched for pairing

$$\mathcal{P}_T^C = (1-\mathcal{P}^{m\to f})\mathcal{P}_O^m + \mathcal{P}^{m\to f}\mathcal{P}_E^u\mathcal{P}_k^f + \mathcal{P}^{m\to f}(1-\mathcal{P}_E^u). \quad (19)$$

- The MU is offloaded as a CEU and a CCU is searched for pairing

$$\mathcal{P}_T^E = (1-\mathcal{P}^{m\to f})\mathcal{P}_O^m + \mathcal{P}^{m\to f}\mathcal{P}_C^u\mathcal{P}_k^f + \mathcal{P}^{m\to f}(1-\mathcal{P}_C^u). \quad (20)$$

Remark 5: The equations in (19) and (20) describes the outage probability for Case II and combines three terms. The first terms in both the equations signifies no offloading from MBS to FBS tier. For Case II, we know apriori whether the OMU is a CEU or CCU and the corresponding PU is searched accordingly which can be observed from the second terms of (19) and (20) which indicates that after offloading, the corresponding PU is searched. The third and the last term indicates the absence of corresponding PU and hence the OMU falls in outage.

2) Total Outage Probability after D2D Cooperation: When the FBS fails to find a corresponding CCU[2] for the MU offloaded as CEU, D2D tier cooperates to serve the OMU by converting the D2D pair to D2D group. The DT of the D2D group uses NOMA and serves the OMU under outage, due to unavailability of corresponding PU, along with its DR. The OMU at the FBS can be treated either as a CCU or CEU in the D2D group, irrespective of how it was offloaded to the FBS. Hence, similar to the NC probability at FBS, to decide whether the OMU is CCU or CEU, we also apply NC probability at the D2D tier to identify whether the cooperation to the OMU is performed as a CCU or CEU, with respect to the DR of the cooperating DT. The total outage probability at the MU offloaded as a CEU after D2D cooperation can be written as follows.

- The OMU is served as a CCU by the cooperating DT

$$\mathcal{P}_E^{CO} = (1-\mathcal{P}^{m\to f})\mathcal{P}_O^m + \mathcal{P}^{m\to f}\mathcal{P}_C^u\mathcal{P}_k^f + \mathcal{P}^{m\to f}(1-\mathcal{P}_C^u)\mathcal{P}_{NC}\left(\mathcal{P}_k^d\right). \quad (21)$$

- The OMU is served as a CEU by the cooperating DT

$$\mathcal{P}_E^{CO} = (1-\mathcal{P}^{m\to f})\mathcal{P}_O^m + \mathcal{P}^{m\to f}\mathcal{P}_C^u\mathcal{P}_k^f + \mathcal{P}^{m\to f}(1-\mathcal{P}_C^u)(1-\mathcal{P}_{NC})\left(\mathcal{P}_k^d\right). \quad (22)$$

Remark 6: The above equation describes the D2D cooperation that takes place due to the absence of corresponding CCU for pairing in (20). The third term in (21) and (22) shows that the OMU in outage, due to the absence of corresponding PU, is served using D2D cooperation as explained in Section II-E.

D. Ergodic Rate Analysis

In this section, we derive the ergodic rate of the MBS tier, FBS tier (without NOMA), FBS tier (with NOMA), and D2D tier (with NOMA). Ergodic rate of the FBS tier (with NOMA) and D2D tier (with NOMA) includes both, a CCU and a CEU.

[2]D2D cooperation is performed only when OMU is CEU and a corresponding CCU is not available.

1) Ergodic Rate of MBS Tier and FBS Tier (without NOMA): The ergodic rate of a typical MU is expressed as follows.

Proposition 6: Using [34], the ergodic rate of a typical MU can be calculated as

$$\mathcal{E}_m = \frac{\alpha_m}{\ln 2}\int_0^\infty \frac{1}{z}\Big(1-\mathcal{L}_{|\tilde{h}_m|^2}(zs_m)\Big)\prod_t \mathcal{L}_{I_t}(z\rho_t^I)e^{-z}\,dz, \quad (23)$$

where α_m is the fraction of total bandwidth allocated to typical MU, $s_m = \rho_m P_m$. $\mathcal{L}_{I_t}(s)$ is the LT of interference from t^{th} tier. $\mathcal{L}_{|\tilde{h}_m|^2}(s)$ is the LT of unordered channel gain of MBS tier and can be calculated as

$$\mathcal{L}_{|\tilde{h}_m|^2}(s) = -\pi\lambda_m\mathcal{Y}_m^2\sum_{n=0}^{N}\frac{b_n^m c_n^m}{c_n^m + s}. \quad (24)$$

Similarly, the ergodic rate of FBS tier (without NOMA) can be calculated. We directly write it as

$$\mathcal{E}_f = \frac{\alpha_f}{\ln 2}\int_0^\infty \frac{1}{z}\Big(1-\mathcal{L}_{|\tilde{h}_f|^2}(zs_f)\Big)\prod_t \mathcal{L}_{I_t}(z\rho_t^I)e^{-z}\,dz, \quad (25)$$

where α_f is the fraction of total bandwidth allocated to typical FU, and $s_f = \rho_f P_f$. $\mathcal{L}_{|\tilde{h}_f|^2}$ denotes the LT of unordered channel gain of FBS tier and can be calculated as

$$\mathcal{L}_{|\tilde{h}_f|^2}(s) = -\frac{1}{\mathcal{Y}_f^2}\sum_{n=0}^{N}\frac{b_n^f c_n^f}{c_n^f + s}. \quad (26)$$

Proof: Please see Appendix F.

2) Ergodic Rate of FBS tier and D2D tier (with NOMA): The ergodic rate of k^{th} user served by an FBS using NOMA is given as follows.

Proposition 7: Ergodic rate at k^{th} user served by an FBS (with NOMA) can be expressed as

$$\mathcal{E}_k^f = \frac{1}{\ln 2}\times\int_0^\infty \frac{1}{z}\left(1-\mathcal{L}_{|h_k^f|^2}(zs_k^f)\right)\prod_{l=k+1}^{M_f}\mathcal{L}_{|h_k^f|^2}(zs_l^f)\times \prod_t \mathcal{L}_{I_t}(z\rho_t^I)e^{-z}\,dz, \quad (27)$$

where $s_k^f = \rho_f P_f a_k$, $s_l^f = \rho_f P_f a_l$, and $\mathcal{L}_{|h_k^f|^2}(s)$ is the LT of ordered channel gain of FBS tier which can be calculated as

$$\mathcal{L}_{|h_k^f|^2}(s) = -\psi_k^f\sum_{r=0}^{M_f-k} f_1^f f_2 \times \sum_{T_k^r} f_3\left(\prod_{n=0}^{N}(b_n^f)^{q_n}\right)\frac{s_f}{s+s_f}, \quad (28)$$

where $s_f = \sum_{n=0}^{N} q_n c_n^f$, $f_1^t = \begin{pmatrix} M_t - k \\ r \end{pmatrix}$, such that $t \in \{f, d\}$, $f_2 = \frac{(-1)^r}{k+r}$, $f_3 = \begin{pmatrix} k+r \\ q_0 \dots q_N \end{pmatrix}$. Similarly, we can write the ergodic rate at k^{th} DR of D2D group as

$$\mathcal{E}_k^d = \frac{1}{\ln 2} \times \int_0^\infty \frac{1}{z}\left(1 - \mathcal{L}_{|h_k^d|^2}(zs_k^d)\right) \prod_{l=k+1}^{M_d} \mathcal{L}_{|h_k^d|^2}(zs_l^d) \times \prod_{t_d} \mathcal{L}_{I_{t_d}}(z\rho_{t_d}) e^{-z}\, dz. \tag{29}$$

where $s_k^d = \rho_d P_d a_k$, $s_l^d = \rho_d P_d a_l$, and $\mathcal{L}_{|h_k^d|^2}(s)$ is the LT of ordered channel gains of D2D group and can be calculated as

$$\mathcal{L}_{|h_k^d|^2}(s) = -\psi_k^d \sum_{r=0}^{M_d-k} f_1^d f_2 \times \sum_{T_k^r} f_3 \left(\prod_{n=0}^{N} (b_n^d)^{q_n}\right) \frac{s_d}{s+s_d}, \tag{30}$$

where $s_d = \sum_{n=0}^{N} q_n c_n^d$.
Proof: Please see Appendix G.

E. Total Ergodic Rate

Similar to Section III-C, we calculate the total ergodic rate for both, after offloading without D2D cooperation and after offloading with D2D cooperation.

1) Total Ergodic Rate after Offloading without D2D cooperation: Total ergodic rate of the MU after offloading as CEU and CCU with respect to the available PU at the FBS, i.e., for Case I, can be written as

- The MU is offloaded to FBS (without NOMA)

$$\mathcal{E}_T = (1 - P^{m\to f})\mathcal{E}_m + P^{m\to f}\mathcal{E}_f. \tag{31}$$

- The MU is offloaded as a CCU at FBS

$$\mathcal{E}_T^C = (1 - \mathcal{P}^{m\to f})\mathcal{E}_m + \mathcal{P}^{m\to f}\mathcal{P}_{NC}\mathcal{E}_k^f. \tag{32}$$

- The MU is offloaded as a CEU at FBS

$$\mathcal{E}_T^E = (1 - \mathcal{P}^{m\to f})\mathcal{E}_m + \mathcal{P}^{m\to f}(1-\mathcal{P}_{NC})\mathcal{E}_k^f. \tag{33}$$

Similarly, we can calculate the total ergodic rate when MU is assumed to be offloaded either as a CEU or CCU at FBS and a corresponding PU is searched for pairing, i.e., for Case II, as

- The MU is offloaded as a CCU and a CEU is searched for pairing

$$\mathcal{E}_T^C = (1 - \mathcal{P}^{m\to f})\mathcal{E}_m + \mathcal{P}^{m\to f}\mathcal{P}_E^u\mathcal{E}_k^f. \tag{34}$$

- The MU is offloaded as a CEU and a CCU is searched for pairing

$$\mathcal{E}_T^E = (1 - \mathcal{P}^{m\to f})\mathcal{E}_m + \mathcal{P}^{m\to f}\mathcal{P}_C^u\mathcal{E}_k^f. \tag{35}$$

2) Total Ergodic Rate after D2D Cooperation: Since, the unavailability of a corresponding PU at FBS leads to outage for the OMU, cooperation from the D2D tier is performed to serve the incoming MU offloaded as a CEU, that fails to find a corresponding CCU for pairing. The OMU can be treated either as a CCU or CEU at the D2D tier, irrespective of how it was offloaded to the FBS. Hence, similar to the NC probability at FBS, we also apply NC probability at the D2D tier to identify whether the cooperation to the OMU is performed as a CCU or CEU, with respect to the DR of the cooperating DT. The total ergodic rate after cooperation can be written as follows

- The OMU is served as a CCU by the cooperating DT

$$\mathcal{E}_E^{CO} = (1 - \mathcal{P}^{m\to f})\mathcal{E}_m + \mathcal{P}^{m\to f}\mathcal{P}_C^u\mathcal{E}_k^f + \mathcal{P}^{m\to f}(1-\mathcal{P}_C^u)\mathcal{P}_{NC}\mathcal{E}_k^d. \tag{36}$$

- The OMU is served as a CEU by the cooperating DT

$$\mathcal{E}_E^{CO} = (1 - \mathcal{P}^{m\to f})\mathcal{E}_m + \mathcal{P}^{m\to f}\mathcal{P}_C^u\mathcal{E}_k^f + \mathcal{P}^{m\to f}(1-\mathcal{P}_C^u)(1-\mathcal{P}_{NC})\mathcal{E}_k^d. \tag{37}$$

Remark 7: The calculation of total ergodic rate follows the same manner as that of the total outage probability in Section III-C. The DT decides whether the cooperation to the OMU is performed as CCU or CEU based on the NC probability as mentioned in the third term of (36) and (37).

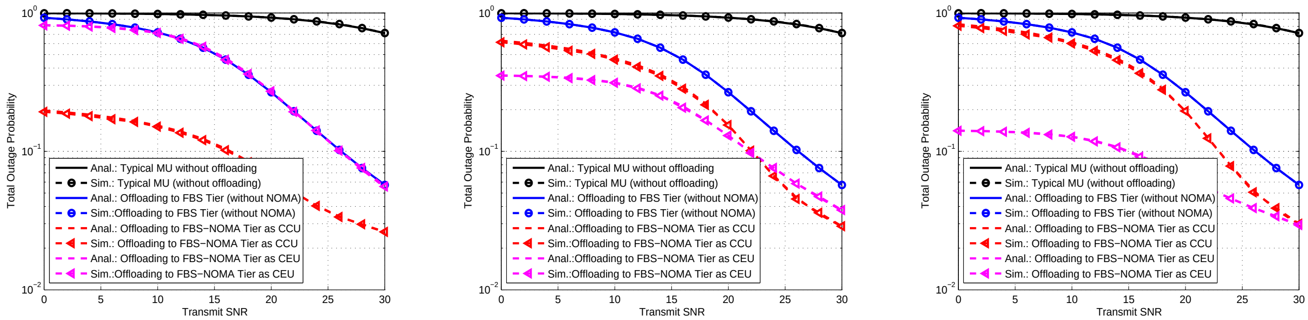
TABLE I: Network Parameters

Symbols	Value
Iterations	10,000
Disk Radius	1000 m
N	10
P_m, P_f, P_d	40 W, 1 W, 3 mW
λ_m, λ_f, λ_d	$5 \times 10^{-5}\text{m}^{-2}, 10^{-4}\text{m}^{-2}, 10^{-4}\text{m}^{-2}$
λ_u	$5 \times 10^{-4}\text{m}^{-2}$
B_m, B_f	$1, 1$
a_k	$0.1, 0.9$
$\mathcal{Y}_m, \mathcal{Y}_f, \mathcal{Y}_d$	1000 m, 5 m, 2 m
α_m, α_f	$0.5, 0.5$
ν_m, ν_f, ν_d	$3, 4, 3$

IV. Results and Discussions

In this section, outage probability and ergodic rate are analyzed based on the analytical expressions derived in Section III for Case I and Case II. The graphs show analytical (Anal.) curves that are verified using Monte Carlo simulation (Sim.) curves. The network parameters for the analytical plots and the simulation plots are given in Table I. We investigate the outage probability and ergodic rate performance for the OMU when D2D tier cooperates in case of non-availability of corresponding PU for the OMU. The impact of cooperation on the FBS tier is also studied using sum ergodic rate.
Note: We refer “FBS-NOMA” for the FBS tier that uses NOMA for transmission. The “FBS tier (without NOMA)” is used to indicate the FBS tier which does not uses NOMA. However, the “FBS tier” mentioned in the discussion would mean that it uses NOMA, unless stated otherwise.

Fig. 2 depicts the variation of outage probability of the typical user of each tier. For the same transmit SNR, D2D tier shows best performance. The coverage range of a DT is small in comparison to the coverage range of MBS and FBS. This means that the distance between the BS and typical user will be smallest for D2D tier and largest for MBS tier. Since, for large distances the path loss is also large, D2D tier shows



(a) For p=0.1.

(b) For p=0.5.

(c) For p=0.8.

Fig. 3: Variation of total outage probability with transmit SNR for different values of p (Case I).

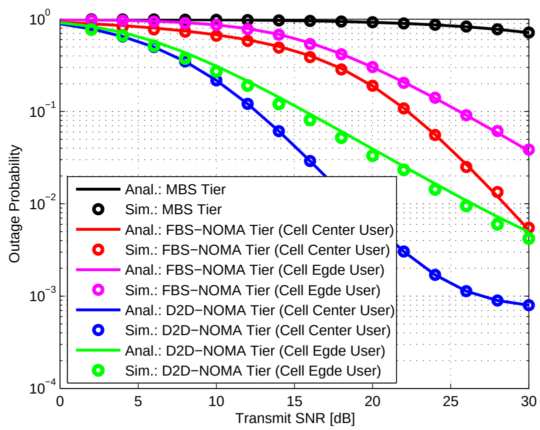
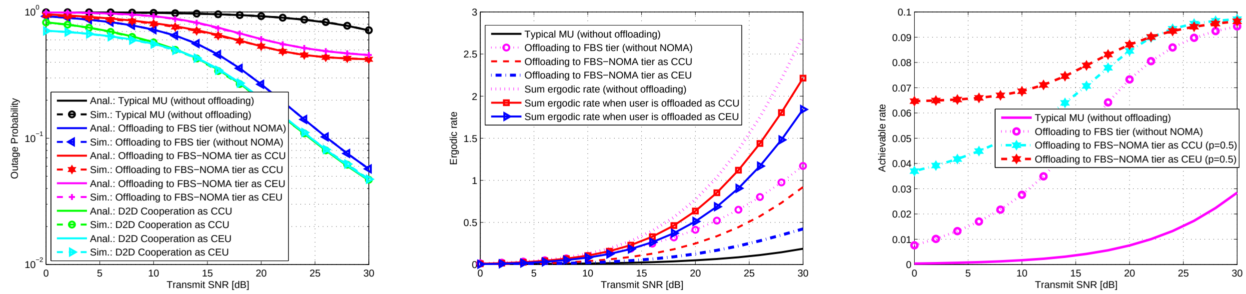


Fig. 2: Variation of outage probability with transmit SNR.

good outage performance as compared to the MBS tier and the FBS tier while the MBS tier shows the worst performance at a given transmit SNR. Also, for the D2D tier, a CCU shows a better outage performance because of better channel condition as compared to the CEU.

Fig. 3 shows the outage probability for different values of p, i.e., for different channel condition of the OMU. It can be clearly observed from all the three graphs of Fig. 3 that offloading to FBS tier (with or without NOMA) gives a better performance at typical MU as compared to without offloading. Fig. 3 reveals the outage performance for Case I when the MU is offloaded to the FBS tier due to congestion at MBS tier. In Fig. 3(a), the analysis is done for $p = 0.1$ which implies that the difference in channel gain between the PU and the OMU is large. As observed from the graph, offloading of the MU to FBS as a CCU yields better performance as compared to if the MU is offloaded to an FBS (without NOMA). However, for the given channel gain difference between the OMU and the PU, offloading of the MU as a CEU gives similar performance as compared to offloading to FBS (without NOMA). Since, p is small, offloading as a CCU would mean a good channel condition for the OMU, whereas, offloading as a CEU gives poor performance owing to its poor channel condition. Hence, for this case, offloading as CEU is nearly equivalent to offloading to FBS (without NOMA), while, offloading as a CCU outperforms offloading to FBS (without NOMA). Fig. 3(b) shows the plot for $p = 0.5$. Since, increase in the value of p indicates decrease in the difference between channel gains of the OMU and the PU, $p = 0.5$ means that the channel gain difference is lesser as compared to $p = 0.1$. Fig. 3(b) shows that for $p = 0.5$ when the MU is offloaded as a CCU or CEU, lower outage probability is achieved as compared to when the MU is offloaded to FBS (without NOMA). It is also observed that the difference in performance of the offloaded CEU and offloaded CCU is smallest for $p = 0.5$ as compared to for $p = 0.1$ and $p = 0.8$. This is due to fact that the power allocation coefficients for CCU and CEU at FBS remains the same, however, the difference in channel conditions is reduced hence, MU offloaded as CEU or as CCU give nearly similar performance. Fig. 3(c) shows the plot for $p = 0.8$. A large value of p indicates that the difference between the channel gain of the OMU and the PU is reduced as compared to $p = 0.1$ and $p = 0.5$. As observed from the graph, when the MU is offloaded as a CEU, lower outage probability is achieved as compared to when MU is offloaded as a CCU. The reason being that the difference in channel gains for this case is less, however the power allocation coefficients remains the same. Thus, a CEU is allocated a higher power allocation coefficient although, the channel difference is not much between CCU and CEU. This means that CEU is served with higher power as compared to CCU although they have a little difference in their channel gains. Hence, outage performance of the MU when offloaded as a CEU is enhanced. Also, for the selected parameters, as given in Table I, whether the MU is offloaded as a CEU or as a CCU, both give better performance as compared to when the MU is offloaded to FBS (without NOMA). To summarize, from Fig. 3 it is clear that if the network setting is such that the OMU is paired with a corresponding PU such that $p = 0.5$, we may say that whether the OMU is a CCU or CEU, it will achieve similar performance and hence, provide user fairness in the proposed HCN-NOMA.

Fig. 4(a) shows outage probability for Case II, when cooperation from D2D tier takes place. For Case I, we assume that the OMU always has a PU for pairing. However, Case II searches for a corresponding user, i.e., if the incoming OMU is a



(a) (b) (c)

Fig. 4: Variation of total outage probability, total ergodic rate and achievable rate with transmit SNR (a) Total outage probability (Case II, $p = 0.5$), (b) Total ergodic rate (Case I, $p = 0.5$), and (c) Achievable rate (Case I).

CEU, a corresponding CCU is searched while, if the incoming offloaded user is a CCU, we search for a corresponding CEU. As observed from Fig. 4(a), the outage performance is deteriorated because the OMU might not find a corresponding PU and fall under outage. Hence, the outage performance for Case II is same, whether the offloaded user is a CCU or CEU and is poor as compared to Case I. Fig. 4(a) also shows the impact of D2D cooperation on the OMU offloaded as a CEU for both the scenarios, i.e., when the the OMU is served as a CCU by the cooperating DT, and when the OMU is served as a CEU by the cooperating DT. When D2D cooperation is performed, whether the OMU is served as a CCU or CEU by the cooperating DT, both show same performance after a transmit SNR of nearly 10dB. For transmit SNR less than 10dB, the OMU served as a CEU by the cooperating DT shows a slightly less outage probability as compared to when the OMU is served as a CCU by the cooperating DT. This is contributed to the less interference observed by the CEU during D2D cooperation, due to poor channel condition. However, for a CCU, the impact of interference is considerable even at low SNR, due to good channel condition. At higher SNR, the negative impact of interference on a CCU is suppressed by additional advantage of increased SNR and both show same performance. With cooperation from D2D tier, Case II shows a decrease in outage probability by $42.76\%, 55.12\%, 73.61\%$, and 86.87% at transmit SNR of 10dB, 16dB, 22dB, and 28dB, respectively, when the offloaded user is a CEU in comparison to without cooperation. Also, D2D cooperation outperforms the offloading to FBS tier without NOMA by a decrease of 21.54% in the outage probability.

Fig. 4(b) shows the ergodic rate and sum ergodic rate for Case I when $p = 0.5$. As observed from Fig. 4(b), when the MU is offloaded as a CCU or CEU, it gives a better performance as compared to no offloading. However, whether the user is offloaded as a CCU or CEU, it gives a poor performance as compared to offloading to FBS tier (without NOMA). Also, the performance of the OMU as a CEU is poor as compared to the OMU as a CCU. Due to poor ergodic rate performance for offloading as a CEU, the sum rate at FBS-NOMA, whether user is offloaded as a CEU or CCU is degraded as compared to sum rate at FBS when no offloading is performed.

Let, Event-E1 denotes that no offloading is performed, and Event-E2 denotes offloading to FBS tier without NOMA. As can be observed from Table II, derived from Fig. 4(b) and Fig. 4(c), increase in the achievable rate is higher than that achieved in terms of sum ergodic rate for Case I, when compared with Event-E1. While comparing with Event-E1, the sum ergodic rate achieves a decrease in performance for Case-I. However, as compared to Event-E2, the sum ergodic rate attains an increase which is slightly higher as compared to that attained in terms of achievable rate. In Case II, this loss in performance in ergodic rate is overcome through cooperation from D2D tier by trying to achieve a higher ergodic rate for user offloaded as a CEU.

Since, D2D cooperation is performed only when the offloaded CEU fails to find corresponding CCU for pairing as explained in Section II-E, hence, Fig. 5 shows all the curves for an offloaded CEU for Case II. Also, whether the OMU is treated as a CEU or CCU in the D2D group, also needs to be taken into account. Hence, we again use NC probability to find whether the OMU is served as a CEU or CCU in D2D group, as done for the MU offloaded at the FBS. The performance graph for D2D cooperation is plotted for $p = 0.5$, for the both the scenarios, i.e., when the OMU is treated as a CCU by the cooperating DT and when the OMU is treated as a CEU by the cooperating DT. As observed from Fig. 5, without cooperation from D2D tier, when the MU is offloaded as a CEU, it achieves a poor performance as compared to when offloading is done to FBS (without NOMA). This is because there are possibilities when the offloaded a CEU does not find a corresponding PU and falls under outage which leads to a degradation in its performance. When cooperation is carried out, in case of absence of corresponding PU, D2D tier helps in serving the offloaded CEU. As can be observed from the graph, D2D cooperation leads to an increase by $2, 2, 1.5$, and 1.4 times in ergodic rate at transmit SNR of 14dB, 22dB, 26dB, and 30dB, respectively, at OMU when served as a CEU by the cooperating DT in comparison to without cooperation. Similarly, D2D cooperation leads to an increase by $3.4, 3, 2.8$, and 2.6 in ergodic rate at transmit SNR of 14dB, 22dB, 26dB, and 30dB, respectively, at the OMU when served as a CCU by the cooperating DT in comparison to without cooperation. The D2D cooperation also outperforms the offloading to FBS

TABLE II: Comparison between achievable rate and sum ergodic rate of Case I with Even-E1 and Event-E2 at SNR= 26dB.

Comparison Increase= New value/Old value	E1 and Case I (when OMU is CCU)	E1 and Case I (when OMU is CEU)	E2 and Case I (when OMU is CCU)	E2 and Case I (when OMU is CEU)
Increase in achievable rate	5.5	5.4	1.06	1.05
Increase in sum ergodic rate	0.82	0.68	1.7	1.6

(without NOMA) with nearly 2 times increase in the ergodic rate (at transmit SNR of 26dB) when the cooperating DT treats the offloaded user as a CCU. However, as compared to offloading to FBS tier (without NOMA) a similar ergodic rate performance is observed when the cooperating DT treats offloaded user as a CEU. Also, there is an increase by $4, 3, 2.7$, and 2.4 times in the sum ergodic rate at FBS-NOMA with D2D cooperation, for transmit SNR of $14\text{dB}, 22\text{dB}, 26\text{dB}$, and 30dB, when the offloaded user is considered as a CEU by the cooperating DT as compared to the sum ergodic rate when no D2D cooperation is performed. Similarly, there is an increase by $4, 2.7, 2.3$, and 2 times in the sum ergodic rate at FBS-NOMA with D2D cooperation, for transmit SNR of $14\text{dB}, 22\text{dB}, 26\text{dB}$, when the offloaded user is considered as a CCU by the cooperating DT as compared to the sum ergodic rate when no D2D cooperation is performed. It can be observed clearly from the above data that the sum ergodic rate, when the offloaded user is treated as a CEU by the cooperating DT, achieves a larger increase as compared to when the offloaded user is treated as a CCU by the cooperating DT. The reason is that the sum ergodic rate consists of sum of ergodic rates of two users and not a single user, and hence, the ergodic rates of the other user in the calculated sum ergodic rate plays a role in this larger increase. Since, the other user is a CCU while calculating the sum ergodic rate when the offloaded user is treated as CEU, it achieves a larger increase as compared to when the offloaded user is treated as CCU, since, the other user in the sum ergodic rate is a CEU. Hence, it can be inferred that cooperation using D2D tier enhances the performance of MU offloaded as CEU. This happens because proposed D2D cooperation gives OMU another chance to get served in case it does not find its corresponding PU. Hence, overall improvement in ergodic rate performance is achieved.

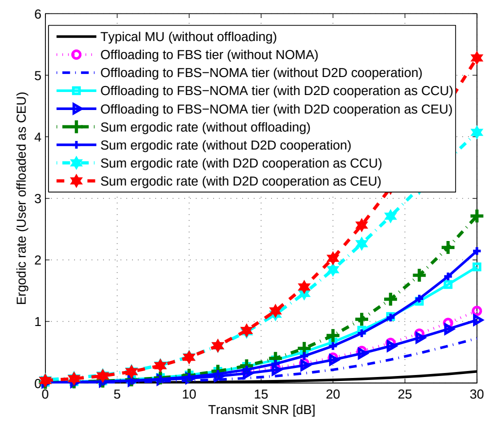


Fig. 5: Variation of total ergodic rates with transmit SNR (Case II, $p = 0.5$.)

V. Conclusion

This work presents a three tier HCN-NOMA framework with different offloading scenarios. The network model incorporates NOMA in FBS tier and in D2D groups. The MU can be offloaded to FBS tier as a CCU or CEU, based on the channel conditions with respect to the available PU with FBS. It is observed that different channel condition during offloading leads to different benefits to the OMU. Also, the offloading MU with the assumption of availability of a PU (Case I) gives better performance as compared to when FBS searches for a corresponding PU according to the OMU (Case II). Since, it is not always possible to find a corresponding PU, the performance is degraded because the OMU goes into outage due to unavailability of corresponding PU. Hence, we have formulated a cooperation scheme using D2D tier, where the D2D pairs transform into D2D groups and use NOMA to serve the offloaded MU under outage along with its own DR. The results show improvement in the outage probability, ergodic rate, and sum ergodic rate performance for the offloaded MU using the proposed D2D cooperation scheme.

Appendix A
Proof of Proposition 1

We have assumed that the typical MU connects to nearest MBS and small scale fading is Rayleigh distributed. With the assumption of homogeneous PPP and applying the polar coordinates, the cumulative density function (CDF) of the unordered channel gain of MBS tier can be written as [8]

$$F_{|\tilde{h}_m|^2}(y) = 2\pi\lambda_m \int_0^{\mathcal{Y}_m} \left(1 - e^{-(1+r_m^{\nu_m})y}\right) e^{-2\pi\lambda_m r_m^2} r_m dr_m. \tag{38}$$

Using G-C quadrature [35], (38) can be approximated as

$$F_{|\tilde{h}_m|^2}(y) \approx \pi\lambda_m \mathcal{Y}_m^2 \sum_{n=0}^{N} b_n^m e^{-c_n^m y}. \tag{39}$$

The outage probability at the typical MU is given as

$$\begin{aligned}\mathcal{P}_O^m &= \mathcal{P}\left(\alpha_m \times \log(1+\text{SINR}_m) < R\right), \quad (40)\\ &= F_{|\tilde{h}_m|^2}\left(\frac{\phi}{\rho_m P_m}\left(1+\sum_t \rho_t^I \mathcal{I}_t\right)\right),\\ &\overset{(a)}{=} \pi\lambda_m \mathcal{Y}_m^2 \sum_{n=0}^{N} b_n^m e^{-c_n^m \frac{\phi}{\rho_m P_m}\left(1+\sum_t \rho_{t_m}^I \mathcal{I}_t\right)},\\ &= \pi\lambda_m \mathcal{Y}_m^2 \sum_{n=0}^{N} b_n^m e^{-c_n^m \frac{\phi}{\rho_m P_m}} \prod_t \mathcal{L}_{\mathcal{I}_t}(s_m \rho_t),\end{aligned}$$

where (a) follows from (39) and α_m is the fraction of bandwidth allocated to typical MU without NOMA.

Appendix B
Proof of Proposition 2

Using the assumption of homogeneous PPP, the CDF of unordered channel gain of FBS tier can be expressed as

$$F_{|\tilde{h}_f|^2}(y) = \frac{2}{\mathcal{Y}_f^2}\int_0^{\mathcal{Y}_f}\left(1-e^{-(1+z^{\nu_f})y}\right) z\, dz. \quad (41)$$

By applying the G-C quadrature [35] to (41), we get

$$F_{|\tilde{h}_f|^2}(y) \approx \frac{1}{\mathcal{Y}_f}\sum_{n=0}^{N} b_n^f e^{-c_n^f y}. \quad (42)$$

The outage probability at the typical FU is given as

$$\begin{aligned}\mathcal{P}_O^f &= \mathcal{P}\left(\alpha_f \times \log(1+\text{SINR}_f) < R\right), \quad (43)\\ &= F_{|\tilde{h}_f|^2}\left(\frac{\phi}{\rho_f P_f}\left(1+\sum_t \rho_t^I \mathcal{I}_t\right)\right),\\ &\overset{(a)}{=} \frac{1}{\mathcal{Y}_f}\sum_{n=0}^{N} b_n^f e^{-c_n^f \frac{\phi}{\rho_f P_f}\left(1+\sum_t \rho_t^I \mathcal{I}_t\right)},\\ &= \frac{1}{\mathcal{Y}_f}\sum_{n=0}^{N} b_n^f e^{-c_n^f \frac{\phi}{\rho_f P_f}} \prod_t \mathcal{L}_{\mathcal{I}_t}(s_f \rho_t^I),\end{aligned}$$

where (a) follows from (42) and α_f is the fraction of bandwidth allocated to typical FU without NOMA.

Appendix C
Proof of Proposition 3

The ordered channel gain of FBS tier has a relationship with the unordered channel gain of FBS tier $F_{|\tilde{h}_f|^2}(y)$ given by [16] as

$$F_{|h_k^f|^2}(y) = \psi_k^f \sum_{r=0}^{M_f-k}\binom{M_f-k}{r}\frac{(-1)^r}{k+r}\left(F_{|\tilde{h}_f|^2}(y)\right)^{r+k}. \quad (44)$$

Substituting (42) in (44) and applying multinomial theorem we get the CDF of ordered channel gain as

$$\begin{aligned}F_{|h_k^f|^2}(y) = \psi_k^f \sum_{r=0}^{M_f-k}\binom{M_f-k}{r}\frac{(-1)^r}{k+r}\times\\ \sum_{T_k^r}\binom{k+r}{q_0\ldots q_N}\left(\prod_{n=0}^{N} b_n^{f\,q_n}\right) e^{-\sum_{n=0}^{N} q_n c_n^f y}. \quad (45)\end{aligned}$$

We derive the outage probability at user k as $\mathcal{P}_k^f = \mathcal{P}\left(\text{SINR}_{k\to j}^f < \phi_j, \text{SINR}_k^f < \phi_k\right)$, $\text{SINR}_{k\to j}$ and SINR_k are given in (3) and (4), respectively. Since, the outage probability is decided on successful SIC followed by successful decoding of self message, we can write outage probability at user k as

$$\mathcal{P}_k^f = \mathcal{P}\left(|h_k^f|^2 < \frac{\epsilon_{max}^f(1+\sum_t \rho_t^I \mathcal{I}_t)}{\rho_f P_f}\right). \quad (46)$$

This gives the outage probability as

$$\mathcal{P}_k^f = F_{|h_k^f|^2}(y), \quad (47)$$

where $y = \frac{\epsilon_{max}^f(1+\sum_t \rho_t^I \mathcal{I}_t)}{\rho_f P_f}$. For a user with $k=1$, outage probability is simply calculated with $y = \frac{\epsilon_j^f(1+\sum_t \rho_t^I \mathcal{I}_t)}{\rho_f P_f}$, since, it decodes only its own message and does not perform SIC. Hence, the outage probability of user k can be calculated using (47) and (45) as expressed in (10). Following the similar procedure, we can write the outage probability at k^{th} DR of the D2D group as expressed in (11)

Appendix D
Proof of Proposition 4

Offloading is based on maximum BRP [19] and a user is associated with the strongest BS in terms of long-term averaged BRP at the user. Hence, the offloading probability can be calculated as follows

$$\begin{aligned}\mathcal{P}^{m\to f} &= \mathbb{E}_{r_f}\left[\mathcal{P}\left(B_m P_m r_m^{-\nu_m} < B_f P_f r_f^{-\nu_f}\right)\right], \quad (48)\\ &= \mathbb{E}_{r_f}\left[\frac{1}{2}\left(e^{-2\pi\lambda_m r_f^{\frac{2\nu_f}{\nu_m}}\left(\frac{B_m P_m}{B_f P_f}\right)^{\frac{2}{\nu_m}}} - e^{-2\pi\lambda_m \mathcal{Y}_m^2}\right)\right].\end{aligned}$$

The probability distribution of r_f can be expressed as $f(r_f) = 2r_f/\mathcal{Y}_f^2$, assuming uniform distribution of FU around FBS within radius $\mathcal{Y}_f$ and r_m follows $f(r_m) = 2\pi r_m \lambda_m \times e^{-\pi r_m^2 \lambda_m}$, owing to NN policy. Hence, we get a closed form expression for the offloading probability as expressed in (12), by using $\nu_m = 3$ and $\nu_f = 4$.

Appendix E
Proof of Proposition 5

When an incoming OMU is received by the FBS, it searches for a corresponding user for pairing. The search is carried out in the CCU region, when the OU is a CEU while the search is carried out in the CEU region, when the OMU is a CCU. Hence, we calculate the probability of finding atleast one user in the desired region of search. To find a CCU for pairing, the region of search is the CCU region, defined by the ESB. Since, we consider the distribution of users to follow PPP process with uniform density λ_u respectively, then the number of users in $\mathcal{Y}_{eq}$ will have a Poisson distribution with mean $\lambda_u \pi \mathcal{Y}_{eq}^2$. Hence, the probability of presence of k users in the ESB can be written as $\mathcal{P}(k) = \frac{e^{-\lambda_u \pi \mathcal{Y}_{eq}^2}(\lambda_u \pi \mathcal{Y}_{eq}^2)^k}{k!}$. Hence, the probability of atleast one user lying in the CCU region is given as

$$\begin{aligned}\mathcal{P}_C^u &= 1-\mathcal{P}(0), \quad (49)\\ &= 1-e^{-\lambda_u \pi \mathcal{Y}_{eq}^2},\\ &= \mathbb{E}_{r_c}\left[1-e^{-\lambda_u \pi \left(\frac{P_f}{P_m}\right)^{\frac{2}{\nu}} r_c^2}\right],\\ &= 1-2\pi\lambda_m \int_o^{\mathcal{Y}_m} r_c e^{-C\pi r_c^2},\\ &= 1-\frac{\left(1-e^{\pi(-C)\mathcal{Y}_m^2}\right)\lambda_m}{C},\end{aligned}$$

where $C = \lambda_u \left(\frac{P_f}{P_m}\right)^{\frac{2}{\nu}} + \lambda_m$. Similar to the number of users in ESB, number of users in the annulus region of FBS coverage ($\mathcal{Y}_f$) and ESB ($\mathcal{Y}_{eq}$) will have a Poisson distribution with mean $\lambda_u\pi(\mathcal{Y}_f^2-\mathcal{Y}_{eq}^2)$. Hence, the probability of presence of k users in the annulus region is given as $\mathcal{P}(k) = \frac{e^{-\lambda_u\pi(\mathcal{Y}_f^2-\mathcal{Y}_{eq}^2)}(\lambda_u\pi(\mathcal{Y}_f^2-\mathcal{Y}_{eq}^2))^k}{k!}$. Therefore, the probability of presence of a PU in CEU region can be written as

$$\begin{aligned}\mathcal{P}_E^u &= 1-\mathcal{P}(0), \quad (50)\\ &= \mathbb{E}_{r_c}\left[1-e^{-\lambda_u \pi (\mathcal{Y}_f^2-\left(\frac{P_f}{P_m}\right)^{\frac{2}{\nu}} r_c^2)}\right],\\ &= 1-\frac{\left(1-e^{\pi(-C)\mathcal{Y}_m^2}\right)e^{-\lambda_u\pi\mathcal{Y}_f^2}\lambda_m}{C}. \quad (51)\end{aligned}$$

Appendix F
Proof of Proposition 6

Using $f_{|\tilde{h}_m|^2}(y) = d(F_{|\tilde{h}_m|^2}(y))/dy$, the PDF of unordered channel gain for MBS tier can be calculated from the CDF of unordered channel gain given in (39), as

$$f_{|\tilde{h}_m|^2}(y) = -\pi\lambda_m\mathcal{Y}_m^2 \sum_{n=0}^{N} b_n^m c_n^m e^{-c_n^m y}. \quad (52)$$

Now, we calculate the LT of unordered channel gain of MBS tier $\mathcal{L}_{|\tilde{h}_m|^2}(s)$. The LT, $\mathcal{L}_X(s)$, can be calculated as $\mathcal{L}_X(s) = \int_0^\infty e^{-sx} f_X(x)dx$, where $f_X(x)$ denotes the PDF of X. Using (52) and the LT formula, we get the LT of unordered channel gain of MBS tier as expressed in (24). The SINR at a typical MU as given in (1) can be rewritten as $\text{SINR}_m = \frac{Y}{\sum_t \rho_t^I \mathcal{I}_t+1}$, where $Y = \rho_m P_m |h_m|^2$. The ergodic rate at the typical MU can be calculated as $\mathcal{E}_m = \alpha_m \times \mathbb{E}\left[\log_2(1+\text{SINR}_m)\right]$. Using [34], we may write the $\ln(1+x)$ term as $\ln(1+x) = \int_0^\infty \frac{1}{z}(1-e^{-xz})e^{-z}dz$. Hence, we solve the ergodic rate as

$$\begin{aligned}\mathcal{E}_m =& \alpha_m \times \mathbb{E}\left[\log_2(1+\text{SINR}_m)\right], \quad (53)\\ \stackrel{(a)}{=}& \frac{\alpha_m}{\ln 2}\mathbb{E}\left[\int_0^\infty \frac{1}{z}\left(1-e^{-z\text{Y}}\right)e^{-z\left(\sum_t \rho_t^I \mathcal{I}_t+1\right)}dz\right],\\ =& \frac{\alpha_m}{\ln 2}\int_0^\infty \frac{1}{z}\left(1-\mathbb{E}\left[e^{-z\text{Y}}\right]\right)\mathbb{E}\left[e^{-z\sum_t \rho_t^I \mathcal{I}_t}\right]e^{-z}dz,\end{aligned}$$

where (a) follows from change in variable as $z = (\sum_t \rho_t^I \mathcal{I}_t+1)z'$ and later by plugging $z' \to z$. Hence, we get the ergodic rate at the typical MU as expressed in (23). Similar to (24), we may calculate the LT of unordered channel gain of FBS tier. Using $f_{|\tilde{h}_f|^2}(y) = d(F_{|\tilde{h}_f|^2}(y))/dy$, the PDF of unordered channel gain for FBS tier can be calculated from the CDF of unordered channel gain given in (42), as

$$f_{|\tilde{h}_f|^2}(y) = -\frac{1}{\mathcal{Y}_f}\sum_{n=0}^{N} c_n^f b_n^f e^{-c_n^f y}. \quad (54)$$

Hence, using (54) and the LT formula, we calculate the LT of unordered channel gain of FBS tier as given in (26). Using (26) and (8), we calculate the ergodic rate at typical FU without NOMA as expressed in (25).

Appendix G
Proof of Proposition 7

Again using $f_{|h_k^f|^2}(y) = d(F_{|h_k^f|^2}(y))/dy$, the PDF of ordered channel gain of FBS tier can be written as

$$\begin{aligned}f_{|h_k^f|^2}(y) = -\psi_k \sum_{r=0}^{M_f-k} f_1^f f_2 \sum_{T_k^r} f_3 \left(\prod_{n=0}^{N}(b_n^f)^{q_n}\right)\times\\ \sum_{n=0}^{N} q_n c_n^f e^{-\sum_{n=0}^{N} q_n c_n y}. \quad (55)\end{aligned}$$

Using (55) and the LT formula, the LT of order statistics $|h_k^f|^2$ can be written as given in (28). SINR at user k to decode its own message as given in (4), can be rewritten as $\text{SINR}_k^f = \frac{Y_k}{Y_l+\sum_t \rho_t^I \mathcal{I}_t+1}$, where $Y_k = \rho_f P_f a_k |h_k|^2$, $Y_l = \rho_f P_f |h_k|^2 \sum_{l=k+1}^{M_f} a_l$.

Following the same procedure as in (53) we get

$$\begin{aligned}\mathcal{E}_k^f &= \mathbb{E}\left[\log_2(1+\text{SINR}_k^f)\right], \quad (56)\\ &= \frac{1}{\ln 2}\mathbb{E}\left[\int_0^\infty \frac{1}{z}\left(1-e^{-zY_k}\right)e^{-z\left(\sum_t \rho_t^I \mathcal{I}_t+Y_l+1\right)}dz\right],\\ &= \frac{1}{\ln 2}\int_0^\infty \frac{1}{z}\left(1-\mathbb{E}\left[e^{-zY_k}\right]\right)\mathbb{E}\left[e^{-zY_l}\right]\mathbb{E}\left[e^{-zY_s}\right]e^{-z}dz,\\ &\stackrel{(a)}{=} \frac{1}{\ln 2}\int_0^\infty \frac{1}{z}\left(1-\mathbb{E}\left[e^{-zY_k}\right]\right)\prod_{l=k+1}^{M_f}\left(\mathbb{E}\left[e^{-zY_l'}\right]\right)\times\\ &\qquad \mathbb{E}\left[e^{-z\sum_t \rho_t^I \mathcal{I}_t}\right]e^{-z}dz,\end{aligned}$$

where (a) follows from the fact that Y_l and $Y_s = \sum_t \rho_t^I \mathcal{I}_t$ are summation terms. $Y_l' = \rho_f P_f a_l |h_k|^2$, thus we can write the final ergodic rate at user k as expressed in (27). Similarly, we can write the ergodic rate at k^{th} DR as expressed in (29). The LT of the ordered channel gain for D2D tier, $\mathcal{L}_{|h_k^d|^2}(s)$, can be calculated following the similar steps as for the FBS tier and is given in (30).

References

[1] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, 2011.

[2] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of k-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commu.*, vol. 30, no. 3, pp. 550–560, 2012.

[3] H.-S. Jo, P. Xia, and J. G. Andrews, "Open, closed, and shared access femtocells in the downlink," *EURASIP J. Wireless Commun. and Netw.*, vol. 2012, no. 1, pp. 1–16, 2012.

[4] W. Bao and B. Liang, "Stochastic analysis of uplink interference in two-tier femtocell networks: Open versus closed access," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6200–6215, 2015.

[5] R. Razavi, M. Dianati, and M. A. Imran, "Non-orthogonal multiple access (NOMA) for future radio access," in *5G Mobile Commun.* Springer, 2017, pp. 135–163.

[6] E. LTE, "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN) (3GPP TS 36.300, version 8.11. 0 release 8), December 2009," *ETSI TS*, vol. 136, no. 300, p. V8, 2011.

[7] E. U. T. R. Access, "Further advancements for E-UTRA physical layer aspects," *3GPP TR 36.814, Tech. Rep.*, 2010.

[8] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, 2014.

[9] Z. Zhang, Z. Ma, M. Xiao, G. Liu, and P. Fan, "Modeling and analysis of non-orthogonal MBMS transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2221–2237, 2017.

[10] Y. Liu, Z. Qin, M. Elkashlan, A. Nallanathan, and J. A. McCann, "Non-orthogonal multiple access in large-scale heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2667–2680, 2017.

[11] J. Zhao, Y. Liu, K. K. Chai, A. Nallanathan, Y. Chen, and Z. Han, "Spectrum allocation and power control for non-orthogonal multiple access in HetNets," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5825–5837, 2017.

[12] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, 2016.

[13] J. He and Z. Tang, "Low-complexity user pairing and power allocation algorithm for 5G cellular network non-orthogonal multiple access," *Electron. Lett.*, vol. 53, no. 9, pp. 626–627, 2017.

[14] P. Swami, M. K. Mishra, and A. Trivedi, "Analysis of downlink power control and cooperation scheme for two-tier heterogeneous cellular network," *International J. Commun. Syst.*, vol. 30, no. 13, p. e3282, 2017.

[15] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, 2016.

[16] Y. Liu, Z. Ding, M. Elkashlan, and J. Yuan, "Non-orthogonal multiple access in large-scale underlay cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10 152–10 157, 2016.

[17] Y. Liu, M. Elkashlan, Z. Ding, and G. K. Karagiannidis, "Fairness of user clustering in MIMO non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1465–1468, 2016.

[18] Y.-G. Lim, C.-B. Chae, and G. Caire, "Performance analysis of massive MIMO for cell-boundary users," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6827–6842, 2015.

[19] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, 2012.

[20] P. Swami, M. K. Mishra, and A. Trivedi, "Performance analysis of two-tier cellular network using power control and cooperation," in *IEEE International Conf. Advances Comput., Commun. Inform. (ICACCI), 2016*, pp. 322–327.

[21] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, 2013.

[22] H. ElSawy, E. Hossain, and M.-S. Alouini, "Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4147–4161, 2014.

[23] D. Zhu, J. Wang, A. L. Swindlehurst, and C. Zhao, "Downlink resource reuse for device-to-device communications underlaying cellular networks," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 531–534, 2014.

[24] K. Doppler, C.-H. Yu, C. B. Ribeiro, and P. Janis, "Mode selection for device-to-device communication underlaying an LTE-advanced network," in *IEEE Wireless Commun. Netw. Conf. (WCNC), 2010*, pp. 1–6.

[25] J. Zhao, Y. Liu, K. K. Chai, Y. Chen, M. Elkashlan, and J. Alonso-Zarate, "NOMA-based D2D communications: Towards 5G," in *IEEE Global Commun. Conf. (GLOBECOM), 2016*, pp. 1–6.

[26] Z. Ding, L. Dai, and H. V. Poor, "MIMO-NOMA design for small packet transmission in the Internet of Things," *IEEE Access*, vol. 4, pp. 1393–1405, 2016.

[27] S. Mukherjee, "Distribution of downlink SINR in heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 575–585, 2012.

[28] J. Venkataraman, M. Haenggi, and O. Collins, "Shot noise models for outage and throughput analyses in wireless ad hoc networks," in *IEEE Military Commun. Conf. (MILCOM) ,2006*, pp. 1–7.

[29] C. Zhai, W. Zhang, and G. Mao, "Cooperative spectrum sharing between cellular and ad-hoc networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 4025–4037, 2014.

[30] G. Mao, B. Fidan, and B. D. Anderson, "Wireless sensor network localization techniques," *Comput. Netw.*, vol. 51, no. 10, pp. 2529–2553, 2007.

[31] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.

[32] K. Subrahmaniam, "On some applications of Mellin transforms to statistics: Dependent random variables," *SIAM J. Appl. Mathematics*, vol. 19, no. 4, pp. 658–662, 1970.

[33] P. Tarasak, S. E. Nai, T. Q. Quek, and F. Chin, "Location-based transmit power control for femtocell access points," in *IEEE ICC, 2012*, pp. 6840–6844.

[34] K. A. Hamdi, "A useful lemma for capacity analysis of fading interference channels," *IEEE Trans. Commun.*, vol. 58, no. 2, 2010.

[35] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*. Springer Sci. Business Media, 2013.